(12) UK Patent Application (19) GB (11) 2 397 139 (13) A

(43) Date of A Publication 14.07.2004

(21) Application No: 0300268.0

(22) Date of Filing: 07.01.2003

(71) Applicant(s):
Intellprop Limited
(Incorporated in the Channel Islands)
PO Box 626, National Westminster House,
Le Truchot, ST PETER PORT, Guernsey,
Channel Islands

(72) Inventor(s):
Jeffrey Wilson

(74) Agent and/or Address for Service:
D Young & Co
21 New Fetter Lane, LONDON, EC4A 1DA,
United Kingdom

(51) INT CL[7]:
G06F 17/60 // G06F 17/28

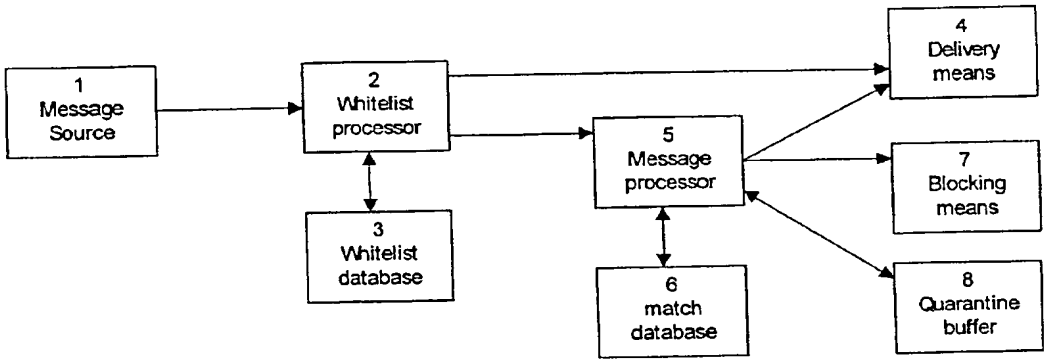
(52) UK CL (Edition W):
G4A AUDB AUDL

(56) Documents Cited:
WO 2001/053965 A1 **JP 2000010880 A**

(58) Field of Search:
UK CL (Edition W) **G4A**
INT CL[7] **G06F, H04L**
Other: **Online: WPI, EPODOC, PAJ**

(54) Abstract Title: **Telecommunications services apparatus for countering spam**

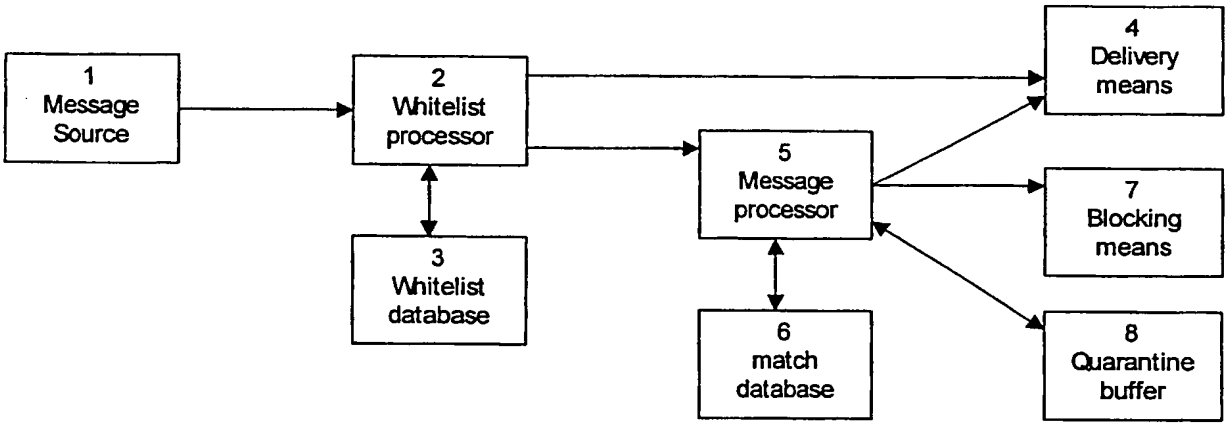
(57) In a text messaging system, occurrences of predefined patterns of characters are identified in a text message, sequences of characters are extracted from positions in the text message relative to the identified patterns, and occurrences of pairings of the extracted sequences and identified patterns are estimated over a recent time window, thereby providing a measure against which message routing or blocking decisions may be made. Counts of detected occurences may be maintained in a database, and corresponding messages may be blocked upon the count reaching a threshold.



GB 2 397 139 A

At least one drawing originally filed was informal and the print reproduced here is taken from a later filed formal copy.

The claims were filed later than the filing date but within the period prescribed by Rule 25(1) of the Patents Rules 1995.



TELECOMMUNICATIONS SERVICES APPARATUS

This invention concerns the field of communications networks and in particular technologies for messaging, including but not limited to email and mobile network text messaging.

As person-to-person communication has become widely used and available through the media of mobile telephony and Internet, so certain aspects of communication, in particular textual communication, have become open to abuse. Where message sending incurs low or zero cost to the sender, bulk messaging may be used to send unsolicited messages to recipients for a variety of purposes. These purposes are predominantly commercial, but are frequently also unlawful, inappropriate for the recipients or are designed to incite the recipient to reply and hence incur charges that benefit the sender and/or the telephone network. In most cases these types of messaging are undesirable and unwanted by recipients. Collectively the unsolicited messages have become known as 'Spam' and in some cases, especially where incitement to reply and thereby inadvertently incur premium rate charges or other unexpected costs, may also be known as 'Scams'. The term Spam is used in this text to cover all such examples.

Many operators of messaging systems have implemented means for detecting and reducing the effects of Spam from their users, with varying degrees of success. However the senders of such Spam have also been able to circumvent many of the blocking measures, and consequently Spam is an ongoing and growing problem in many messaging domains.

In mobile telephony, text messaging including SMS, EMS, MMS and the like, has been affected by Spam, and despite guidelines and laws within some recipients' territories, the global nature of mobile telephony has made it difficult for operators to prevent Spam from arriving from sources outside their own networks.

The present invention provides a relatively simple but effective scheme for the detection and blocking of Spam messages, that is applicable to all kinds of textual communication media, but is of particular application in mobile telephony and Internet email scenarios where automated detection and blocking is required due to the enormous volume of messaging to be filtered.

The invention is described here in the context of a GSM mobile telephone network, although the invention is applicable to other types of mobile network and to email networks.

Known techniques may be used for routing text message traffic via equipment such as an SMS Router that is capable of analysing message content and making routing or blocking decisions accordingly. Similarly for Internet networks, known techniques may be employed for routing all messages to a filtering mechanism prior to delivery to recipients. This patent is concerned with certain processing techniques that may be employed within equipment that has access to messages to be filtered.

According to the invention there is provided a Telecommunications Services Apparatus, the apparatus comprising means for identifying the occurrence of predefined patterns of characters in a message, means for extracting sequences of characters from positions in the message relative to the identified patterns, means for estimating the relative frequency of occurrence of the extracted sequences over a recent time window, thereby providing a measure against which routing or blocking decisions may be made on a per-message basis.

Messages are delivered to the apparatus from message sources (1) by known means that are outside the scope of this invention. A white list processor (2) identifies by means of a white list database (3) messages that are from known or trusted sources that may be delivered via message delivery means (4) without Spam filtering. Messages to be filtered are passed to the message processor (5) which in conjunction with a match database (6) identifies messages to be blocked. Blocked messages are sent to the message blocking means (7) or discarded whilst other messages are delivered via

message delivery means (4). Suspect messages may be temporarily quarantined in a quarantine buffer (8).

In the GSM mobile telephony system, SMS text messages support person-to-person and machine to person messages of up to 160 characters, and longer messages are possible if message concatenation is used. It is desirable to apply automatic processing to these messages such that Spam is selectively blocked, without disrupting genuine person-to-person messaging, or genuine bulk Host-to person subscription messaging, for example "Goal!" alerts.

In a preferred embodiment, an SMS Router is used, which is capable of utilising addressing information as well as message content for performing filtering operations and for making routing or blocking decisions. The SMS Router may also be connected to an SMS SCP (Service Control Point) to increase the flexibility and power of its filtering capability.

An additional difficulty is that what is a desired message to one person may be Spam to another. For example, networks themselves are increasingly sending one or more 'welcome messages' to roamers when they log-on to the roaming network. These messages are typically advertising short codes that may be used on the roaming network to access premium services, or numbers for operator assistance. While useful to a minority of users, many regard these messages as Spam. Therefore no solution can provide perfect automated Spam rejection, since the definition of Spam varies between subscribers. However it is possible to provide blocking for classes of messages filtered according to operator-selected criteria, chosen to suit the majority of subscribers.

The characteristics of person-to-person messaging are such that a maximum message rate is practically defined by the minimum time it takes for a person to enter and send a new message. However, the possibility of group messaging must also be taken into account, whereby a user may send the same message to multiple recipients, by using a feature of the network or his handset, therefore message sending rate alone cannot be used as a criterion for identifying Spam.

Valid use of bulk messaging by SMS Hosts to subscribers must not be affected, therefore repeated use of the same message content cannot be used alone as a criterion either. However it is possible to white list allowed CLIs (numeric or alphanumeric) or the originating global titles (set by the network) for trusted sending equipments. These may be solely within the operator's network, or may also include trusted service providers from other networks. These addresses are then allowed to send messages without being filtered by the Spam equipment. A disadvantage of this approach is that every service provider that is to be allowed to send bulk messaging to a subscriber has first to be enabled in the operator's white list. However since genuine bulk messaging is essentially indistinguishable from Spam except by means of the sender's identity, this type of approach has to be taken if Spam is to be automatically filtered out.

The proposed algorithm is now described, assuming that all messages directed to the apparatus are to be filtered according to the algorithm, and that messages from trusted sources have been directed away from the apparatus, or otherwise prevented from being filtered.

The proposed algorithm works by counting the number of occurrences of sequences of characters called *signatures*. Signatures are defined as string fields at a fixed position relative to *anchors*. An anchor is either

- one of a set of predefined text patterns or sequences that may occur anywhere in the message or
- some other text specification, for example the second word, or the longest word, or a premium rate phone number, which is recognisable by prefix and length.

Each anchor also defines the position of one or more signature fields relative to itself. For example signatures could be preferably be defined as follows:

signature = the text between each anchor and the next anchor, taking the end of the message as an implicit anchor.

Prior to processing, the text may be pre-processed by rationalising case, punctuation and white space. For example, certain punctuation could be removed (e.g. apostrophes) all non-printing characters and certain symbols and other punctuation characters could be replaced by a single space, and then any occurrences of multiple spaces could be reduced to a single space. Case could be converted to, for example, all lower case to remove the variability of jittered case that may be used by a Spam attacker to render the message harder to match.

Attackers may also use a range of telephone numbers in Spam messages rather than a single number in order to evade exact match algorithms. For this reason anchors utilising number strings preferably will only make use of the number prefix rather than a complete number, and support wildcarding for the remainder of the number so as to prevent the variable part of the number from generating a range of signatures. For example, an anchor specification such as +34123456* would match against +3412345678, +3412345600 and +341234567890123 and the variability in these numbers would not be reflected in the adjacent signatures that were extracted.

These compressions and wildcarding techniques reduce the variability in the message, especially types of deliberate variability that can be introduced by Spammers while retaining readability of the message.

Next, the anchors are searched for. This may be done by comparing words from a list of anchor words or specifications with the text to find matches. From each anchor the signature strings are extracted. These are found relative to the anchors according to the anchor definition. Finally the extracted signatures are hashed into a table or database. The database is indexed by a hashed version of the anchor definition. Against each indexed is stored a list of hashed signatures, each with a count and a timestamp. Occurrences of the same anchor/signature pairing will cause a counter for that hashed table entry to be incremented. Each item in the list has its own counter.

The aim is to use the anchor words to identify signatures, which in the case of Spam will be duplicated in numerous messages from the same source. If we automatically

identify and block messages with identical signatures when their occurrence exceeds a predefined threshold, we can largely eliminate Spam. By forcing Spam publishers to continually change their text, we make it hard to automatically generate.

Example:

If the strings casino, day holiday, mystery and send msg are defined as anchor words, signatures are defined as the text following each anchor up to the start of the next anchor (or the end of the message). An example text message is:

Great offer. Casino prizes. Free 14 day holiday for two in the Seychelles or mystery musical prize. Send msg to +341234567890. Max cost £3

Then, the signatures are

- prizes. Free 14
- for two in the Seychelles or
- musical prize.
- to +34123456789. Max cost £3

Each of these signatures is then hashed into the database list associated with each anchor, i.e. 'prizes. Free 14' is hashed into the list indexed by the hashed anchor 'casino'. If a corresponding hashed value already exists in the list, then that list item's counter is incremented, otherwise a new list item is created and its counter is set to 1. If the list now exceeds a pre-defined maximum length, then an item is deleted from the list. The item to delete is preferably chosen according to a combination of count and timestamp, such that a lower count and an older timestamp make deletion more likely.

The effect of this algorithm is to maintain a database of hashed occurrences of strings that occur between defined anchors. For a count to be incremented the corresponding anchor and signature pairing must recur. To resist deletion and replacement by other pairings, a list entry must have a relatively young creation timestamp, and a sufficiently high count. These characteristics exactly mirror the characteristics of Spam, in that a large number of occurrences of signature strings are transmitted in a relatively short period.

Hence the database may be used to form a decision criterion for routing or blocking a message. If an anchor/signature pair is found in a message, and the database has a matching hashed entry with a count that exceeds a pre-defined threshold, then this indicates that this anchor signature pair has started occurring relatively recently and has occurred a given number of times since. Therefore the messages may be marked as Spam, and treated accordingly.

With this technique Spam is blocked once the decision threshold has been reached, however an improvement to the algorithm can allow occurrences of Spam to be blocked prior to the threshold being reached. To do this a delay may be introduced using a quarantine buffer (8) such that after processing by the Spam detection algorithm, messages that exhibit certain anchor words or combinations of anchor words but do not exceed the Spam detection threshold, are tentatively tagged as Spam and routed to the quarantine buffer. This buffer is arranged as a first-in first out buffer with a nominal delay that is set at a level that is typically long enough for Spam attacks to have been adequately registered by the described algorithms so as to permit successful detection of further occurrences. Messages exiting the quarantine buffer after the delay are then checked again against the Spam detection thresholds, but without modifying the counters. If a quarantined message contains a signature matching a database entry with a count higher than the detection threshold, then the message is identified as Spam and handled accordingly, otherwise the message is delivered normally, having been delayed for a short time.

A key feature of the algorithm is that for a chosen set of anchors, the lists and the database are self-maintaining. This is achieved by defining a maximum length for the list associated with each anchor, and allowing the content of the lists to evolve automatically as signatures are detected and processed.

The list of anchors in use may by modified by the operator in the light of reported occurrences of Spam or occurrences detected by other means.

By efficiently checking for signatures, Spam publishers are forced to continually edit their messages, which will be very expensive, and ultimately futile.

Anchor words are predefined, and the list may be continually modified or added to over time to improve the performance of the algorithm.

The proposed method is to use hashing, which is a known technique for database storage, as an efficient way to compare strings. A hashing process $H = h(x)$ is defined to operate on an arbitrary string x and produce a hash value H, which in the preferred embodiment is an integer of defined maximum length. The maximum length bounds the number of possible hash values H that can be generated, independently of the size of the set of possible strings x, and the choice of an integer makes sorting and comparisons computationally efficient.

The hashing technique is now described in more detail. Hashing is a technique whereby a set of input data which may have a large number of possible variations is mapped to a smaller bounded set of possibilities. As a simple illustrative example, take the set of 11-digit telephone numbers as the input. As each digit has ten possibilities, there are 10^{11} (one hundred thousand million) possible telephone numbers that can be constructed with 11 digits. It would generally be impractical and extremely inefficient for an application that, for example, needed to index people by telephone number to maintain a database of 10^{11} entries, especially if the application related to only a small proportion of the subscriber base, say 10,000 people. In this example, only one in

every 10 million possible numbers would actually be used. The solution is to use hashing. Each of the ten thousand 11-digit numbers that is to be used is hashed to produce an index that is bounded for example to the range 0 to 65535, i.e. a 16 bit value. A database of 65536 entries is then usable to store the required data about each used telephone number. The hashing process could take one of a variety of forms, and may be very simple; for example—

- divide the telephone number by 65536 and take the remainder, or
- multiply each of the digits by a different prime number, add the results together, divide the result by 65536 and take the remainder
- etc.

Each possible hashing function has the property that the result is bounded to the chosen range, and may therefore be used as an index to a database of that size.

Clearly, two different telephone numbers may hash to the same value, and this is called a collision. This problem is addressed by allowing each index in the database to reference multiple entries that are normally stored as a linked list (a collision list). A linked list allows variable length in an efficient manner. Each item in a collision list normally contains a copy of the original (unhashed) data, so that the appropriate entry may be identified.

The operation of the database is now described in more detail. A database is constructed which is indexed by a hashed representation of the set of defined anchors. New anchors may be added, and anchors removed at will. Against each hashed anchor is stored a collision list of zero or more entries. Each collision list entry is identified by a copy of the unhashed anchor, and itself contains a list (a signature list) of zero or more hashed signatures that have occurred following that anchor, and each element of the list has a count value and a timestamp representation that may be used as a coarse indication of how old the list element is. The timestamp is preferably quantised to 1-second resolution, so that a 16 bit value can span about 18 hours. Other resolutions are possible.

The signature lists preferably have a maximum length. As each message is processed, hashed signatures are generated for each anchor in the message, and these are compared against the signature list for the corresponding anchor in the database. If an identical hashed signature value is found in the signature list, then the corresponding counter is increased. If the hashed signature value is not already in the list for this anchor and the list is not already at maximum length, then the hashed signature value is added to the list, with a count of 1 and the current timestamp.

If the signature list for this anchor was already full then an element is deleted before the new value is added. The element to be deleted may be preferably chosen on the basis of some combination of lowest count and oldest timestamp. In this way infrequently occurring and older signatures will tend to be deleted from the list in preference to more frequent and more recent signatures. Count values may also be automatically decayed over time, so that signatures that are no longer occurring will eventually be deleted from the lists also.

When a Spamming event is processed, many messages with identical signatures will be seen in a short period, and these will lead to certain count values in the database rising sharply. A threshold may be set above which messages with these signatures can be identified as Spam and automatically blocked or rerouted.

To prevent the filtering process from identifying genuine Host originated bulk traffic as Spam, certain trusted sources may be white listed according to their originating global title, and allowed to send without passing through the Spam filter.

The database may preferably be arranged to be segregated by originating global title. This means that different network source routing addresses (that cannot be spoofed by the Spam originators in the way the CLI for example could) each have their own logical database and are counted separately from other source addresses. This reduces the likelihood that genuine person-to-person traffic, that is spread over many SMSCs would able to increase counts above the preset threshold for detecting Spam, whilst

bulk traffic from a single foreign SMSC would tend to increase the counts strongly if the messages contained identical signatures.

CLAIMS

1. A telecommunications services apparatus for use in a text messaging system, the apparatus comprising means for identifying occurrence of predefined patterns of characters in a message, means for extracting sequences of characters from positions in the message relative to the identified patterns, and means for estimating occurrences of pairings of the extracted sequences and identified patterns over a recent time window, thereby providing a measure against which routing or blocking decisions may be made for each received message.

2. Apparatus according to claim 1, including a match database operable to identify messages to be blocked.

3. Apparatus according to claim 1 or claim 2, including whitelist means allowing messages identified as being from permitted sources to bypass the identifying, extracting and estimating means, such that block decisions are not made on such messages from permitted sources.

4. Apparatus according to any one of claims 1 to 3, wherein the predefined patterns of characters include alphanumeric characters.

5. Apparatus according to any one of claims 1 to 4, wherein the sequences of characters include alphanumeric characters.

6. Apparatus according to any one of claims 1 to 5, wherein the predefined patterns of characters include numeric strings.

7. Apparatus according to any one of claims 1 to 6, wherein the sequences of characters include numeric strings.

8. Apparatus according to any one of claims 1 to 7, wherein the identifying means includes means listing a plurality of predefined patterns of characters for comparison with received messages.

9. Apparatus according to any one of claims 1 to 8, including hashing means for hashing the extracted sequences of characters into a database.

10. Apparatus according to any one of claims 1 to 9, including counting means for maintaining a count of the extracted sequences of characters, the counting means being incremented each time the same predefined pattern of characters and the same extracted sequence of characters are identified.

11. Apparatus according to claim 10, including means for discarding count information from the counting means, based on a combination of age and count value.

12. Apparatus according to any one of claims 1 to 11, including timing means for providing an indication of the timing of each occurrence of the extracted sequence of characters.

13. Apparatus according to any one of claims 1 to 12, including text pre-processing means for pre-processing text messages before being presented to the identifying means, the pre-processing including at least one of rationalising case, punctuation and white space within the text messages.

14. A telecommunications services apparatus for use in a text messaging system, the apparatus being substantially as herein described with reference to and as illustrated in the accompanying drawing.

15. A telecommunications services method for a text messaging system, the method comprising identifying occurrence of predefined patterns of characters in a message, extracting sequences of characters from positions in the message relative to the identified patterns, and estimating occurrences of pairings of the extracted

sequences and identified patterns over a recent time window, thereby providing a measure against which routing or blocking decisions may be made for each received message.

16. A method according to claim 15, including identifying messages to be blocked by the use of a match database.

17. A method according to claim 15 or claim 16, including whitelisting allowing messages identified as being from permitted sources to bypass the identifying, extracting and estimating steps, such that block decisions are not made on such messages from permitted sources.

18. A method according to any one of claims 15 to 17, wherein the predefined patterns of characters include alphanumeric characters.

19. A method according to any one of claims 15 to 18, wherein the sequences of characters include alphanumeric characters.

20. A method according to any one of claims 15 to 19, wherein the predefined patterns of characters include numeric strings.

21. A method according to any one of claims 15 to 20, wherein the sequences of characters include numeric strings.

22. A method according to any one of claims 15 to 21, wherein the identifying step utilises means listing a plurality of predefined patterns of characters for comparison with received messages.

23. A method according to any one of claims 15 to 22, including hashing the extracted sequences of characters into a database.

24. A method according to any one of claims 15 to 23, including maintaining a count of the extracted sequences of characters, the count being incremented each time the same predefined pattern of characters and the same extracted sequence of characters are identified.

25. A method according to claim 24, including discarding count information based on a combination of age and count value.

26. A method according to any one of claims 15 to 25, including providing an indication of the timing of each occurrence of the extracted sequence of characters.

27. A method according to any one of claims 15 to 26, including pre-processing text messages before being identified, the pre-processing including at least one of rationalising case, punctuation and white space within the text messages.

28. A telecommunications services method for a text messaging system, the method being substantially as herein described with reference to and as illustrated in the accompanying drawing.

The Patent Office
PATENTS • DESIGNS • COPYRIGHT • TRADE MARKS
INVESTOR IN PEOPLE

Application No: GB 0300268.0 **Examiner:** Ben Buchanan
Claims searched: all **Date of search:** 17 March 2004

Patents Act 1977 : Search Report under Section 17

Documents considered to be relevant:

Category	Relevant to claims	Identity of document and passage or figure of particular relevance	
X	1 & 15 at least	WO 01/53965	(ODYSSEY) see whole document
A		JP 2000010880	(RICOH)

Categories:

X	Document indicating lack of novelty or inventive step	A	Document indicating technological background and/or state of the art.
Y	Document indicating lack of inventive step if combined with one or more other documents of same category.	P	Document published on or after the declared priority date but before the filing date of this invention.
&	Member of the same patent family	E	Patent document published on or after, but with priority date earlier than, the filing date of this application.

Field of Search:

Search of GB, EP, WO & US patent documents classified in the following areas of the UKC[W]:

G4A

Worldwide search of patent documents classified in the following areas of the IPC[7]:

G06F, H04L

The following online and other databases have been used in the preparation of this search report:

WPI, EPODOC, PAJ